

**PENDETEKSIAN *INFLUENTIAL OBSERVATION*
PADA MODEL REGRESI LINIER MULTIVARIAT
MENGUNAKAN JARAK COOK TERGENERALISASI
(STUDI KASUS INDIKATOR PENDIDIKAN PROVINSI JAWA TENGAH TAHUN 2010)**

Puti Cresti Ekacitta¹, Diah Safitri², Triastuti Wuryandari³

¹Mahasiswa Jurusan Statistika FSM Universitas Diponegoro

^{2,3}Staf Pengajar Jurusan Statistika FSM Universitas Diponegoro

ABSTRACT

Multivariate linear regression model is regression model with one or more response variable and one or more predictor variable, with each response variable are mutually. In multivariate linear regression model sometimes often found Influential Observation. Influential Observation give most contributing in estimating regression coefficient. For detection Influential Observation on multivariate linear regression model is used Generalized Cook's Distance. The aim of this research is to detection any or not any Influential Observation on multivariate linear regression model of education indicator in Central Java Province with response variable are Gross Participation Rate (APK), School Participation Rate (APS), and Pure Participation Number (APM) and predictor variable is percentage of population aged 10 years and over who graduated from junior high school. Result from this research can be explained that if the percentage of population aged 10 years and over who graduated from junior high school increase one percent, it will have an impact on increasing gross participation rate the junior high school is 1.7849 % , increasing school participation rate is 1.6275 % and increasing pure participation number is 1.3712 %. Also, from this results were obtained two observations are included Influential observation. Elimination of the two observations are included Influential observation in the multivariate linear regression model of education indicators in Central Java, affects the regression coefficients change only and does not have a major impact on the closeness of the relationship between response variables and predictor variables in the multivariate.

Keyword : Multivariate linear regression model, Influential Observation, generalized cook's distance, APK, APS, APM

1. PENDAHULUAN

Pendidikan merupakan hak asasi manusia dan hak setiap warga untuk dapat mengembangkan potensi diri melalui proses pembelajaran maksimal pada tingkat dasar. Hak untuk mendapatkan pendidikan sebagai pemenuhan hak asasi manusia merupakan landasan diselenggarakannya Pendidikan Untuk Semua (PUS). Dalam upaya memperluas pendidikan untuk semua, pemerintah menyelenggarakan pendidikan melalui jalur formal, non formal, dan informal. Berkaitan dengan pemerataan pendidikan, penuntasan wajib belajar untuk semua, menghapus kesenjangan gender, dan meningkatkan mutu pendidikan, pemerintah Indonesia telah mencanangkan gerakan wajib belajar dari enam tahun yang dimulai pada tahun 1984 menjadi sembilan tahun yang dimulai pada tahun 1994 (Badan Pusat Statistik Jawa Tengah, 2010).

Pada tulisan ini dilakukan suatu penelitian mengenai indikator pendidikan di Provinsi Jawa Tengah berdasarkan Kabupaten dan Kota menurut jenjang pendidikan SMP tahun 2010 untuk melihat pencapaian keberhasilan program wajib belajar sembilan tahun, berdasarkan data sekunder dari BPS Provinsi Jawa Tengah. Penelitian ini menggunakan regresi linier multivariat karena terdapat lebih dari satu variabel respon dan satu atau lebih variabel prediktor dengan masing - masing variabel respon saling berhubungan satu dengan yang lain. Adapun variabel yang digunakan sebagai variabel respon adalah Angka Partisipasi Kasar (APK), Angka Partisipasi Sekolah (APS), dan Angka Partisipasi Murni (APM) jenjang pendidikan SMP dan variabel respon adalah persentase penduduk berusia 10 tahun ke atas yang menamatkan jenjang pendidikan SMP.

Selanjutnya akan dideteksi apakah terdapat pengamatan yang merupakan *Influential Observation* pada model regresi linier multivariat untuk indikator pendidikan di Jawa Tengah dan melakukan penanganan jika terdapat *Influential Observation*. *Influential Observation* merupakan suatu pengamatan yang dapat menyebabkan perubahan signifikan pada kesesuaian fungsi regresi (Altunkaynak and Ekni, 2002). Berbagai macam pengukuran untuk mendeteksi *Influential Observation* pada model regresi linier multivariat telah dikembangkan, seperti jarak Welsch-Kuh tergeneralisasi digunakan untuk mengukur adanya pengaruh (*influence*) pada variabel respon, jarak *covariance ratio* digunakan untuk mengukur pengaruh (*influence*)

pada nilai matriks kovarian $\hat{\beta}$, dan jarak Cook tergeneralisasi digunakan untuk mengukur pengaruh (*influence*) pada koefisien regresi termasuk intersep (Timm, 2002). Suatu pengamatan ke-*i* dapat dikatakan terdapat *Influential Observation* jika nilai jarak Cook tergeneralisasi (DC_i) lebih besar dari 1. Setelah terdapat pengamatan yang termasuk *Influential Observation*, penghapusan data dilakukan untuk menilai perubahan pada koefisien regresi terutama estimasi parameter (Altunkaynak and Ekni, 2002).

2. TINJAUAN PUSTAKA

2.1 Pendidikan

Dalam pengertian umum, pendidikan adalah suatu proses untuk meningkatkan pengetahuan guna menjadikan manusia berkualitas dan merupakan kebutuhan dasar untuk meningkatkan kecerdasan dan keterampilan.

Indikator Pendidikan adalah :

a. Angka Partisipasi Kasar (APK)

Angka Partisipasi Kasar merupakan persentase jumlah penduduk yang sedang bersekolah pada suatu jenjang pendidikan terhadap jumlah penduduk usia sekolah yang sesuai dengan jenjang pendidikan tersebut.

b. Angka Partisipasi Sekolah (APS)

Angka Partisipasi Sekolah menunjukkan partisipasi sekolah penduduk khususnya anak usia sekolah dalam proses kegiatan sekolah.

c. Angka Partisipasi Murni (APM)

Angka Partisipasi Murni menunjukkan partisipasi sekolah penduduk dengan memperhatikan kesesuaian umur dan jenjang pendidikan yang diduduki.

(Badan Pusat Statistik Jawa Tengah, 2010)

2.2 Distribusi Normal Multivariat

Distribusi normal multivariat merupakan perluasan dari distribusi normal univariat. Densitas normal multivariat p dimensi untuk vektor random $\mathbf{X}' = [X_1, X_2, \dots, X_p]$ adalah:

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{\frac{p}{2}} |\Sigma|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} (\mathbf{x} - \mu)' \Sigma^{-1} (\mathbf{x} - \mu) \right]$$

dengan $-\infty < x_j < \infty$, $j = 1, 2, \dots, p$, diberi notasi $N_p(\mu, \Sigma)$.

(Johnson and Wichern, 2007)

2.3 Koefisien Korelasi

Koefisien korelasi dirumuskan sebagai berikut :

$$r = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}}$$

Nilai korelasi berkisar antara -1 sampai 1 ($-1 \leq r_{xy} \leq 1$), artinya semakin mendekati 1, berarti hubungan dua variabel tersebut semakin erat secara linier dan juga sebaliknya. Apabila nilai korelasi 0, maka hubungan dua variabel tersebut berarti lemah secara linier atau memungkinkan hubungannya tidak secara linier.

(Johnson and Wichern, 2007)

2.4 Regresi Linier Multivariat

Model regresi linier multivariat adalah model regresi dengan lebih dari satu variabel respon yang saling berkorelasi dan satu atau lebih variabel prediktor. Misal terdapat variabel respon berjumlah q yaitu Y_1, Y_2, \dots, Y_q dan p variabel prediktor yaitu X_1, X_2, \dots, X_p , maka model linier multivariat respon ke- q adalah :

$$\begin{aligned}
Y_1 &= \beta_{01} + \beta_{11}X_1 + \dots + \beta_{p1}X_p + \varepsilon_1 \\
Y_2 &= \beta_{02} + \beta_{12}X_1 + \dots + \beta_{p2}X_p + \varepsilon_2 \\
&\vdots \\
Y_q &= \beta_{0q} + \beta_{1q}X_1 + \dots + \beta_{pq}X_p + \varepsilon_q
\end{aligned}$$

Model regresi linier multivariat yang terdiri dari q model linier secara simultan dapat ditunjukkan dalam bentuk matriks sebagai berikut :

$$\underset{(n \times q)}{\mathbf{Y}} = \underset{(n \times (p+1))}{\mathbf{X}} \underset{((p+1) \times q)}{\boldsymbol{\beta}} + \underset{(n \times q)}{\mathbf{E}}$$

(Johnson and Wichern, 2007)

Estimasi kuadrat terkecil untuk $\boldsymbol{\beta}$ adalah : $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$

(Rencher, 2002)

Pengujian regresi linier multivariat secara keseluruhan dilakukan untuk melihat apakah semua variabel prediktor mempengaruhi variabel respon secara multivariat. Hipotesis yang digunakan sebagai berikut :

$$H_0 : \beta_{11} = \beta_{12} = \dots = \beta_{p1} = \dots = \beta_{pq} = 0$$

$$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$$

Statistik uji yang digunakan adalah *Wilk's lambda* :

$$\Lambda = \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'|} \quad (1)$$

dengan $\bar{\mathbf{y}}$ adalah vektor rata-rata dari matriks \mathbf{Y}

H_0 ditolak jika $\Lambda \leq \Lambda_{\alpha, q, p, n-p-1}$ adalah nilai kritis untuk *Wilk's Lambda*.

(Rencher, 2002)

Pengujian signifikansi parameter dilakukan untuk melihat apakah variabel prediktor secara individu mempengaruhi variabel respon secara multivariat. Hipotesis yang digunakan sebagai berikut :

$$H_0 : \beta_{11} = \beta_{21} = \dots = \beta_{p1} = \dots = \beta_{pq} = 0$$

$$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$$

Statistik uji untuk signifikansi parameter sebagai berikut :

$$\begin{aligned}
\Lambda_x &= \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'_r\mathbf{X}'_r\mathbf{Y}|} \\
&= \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'_r\mathbf{X}'_r\mathbf{Y}|} \frac{|\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'|}{|\mathbf{Y}'\mathbf{Y} - n\bar{\mathbf{y}}\bar{\mathbf{y}}'|} = \frac{\Lambda_f}{\Lambda_r}
\end{aligned}$$

H_0 ditolak jika $\Lambda_x \leq \Lambda_{\alpha, q, h, n-p-1}$ adalah nilai kritis untuk *Wilk's Lambda*, dengan h adalah dengan h adalah banyaknya parameter pada $\boldsymbol{\beta}_d$, d menunjukkan subset pada β_{pq} yang akan dihapus jika variabel prediktor tidak signifikan terhadap variabel respon.

(Rencher, 2002)

Ukuran yang digunakan untuk mengukur hubungan keeratan antara variabel respon dan variabel prediktor pada regresi multivariat adalah dengan menggunakan *Wilk's lambda*. Ukuran ini dinyatakan dengan rumus :

$$\eta_{\Lambda}^2 = 1 - \Lambda$$

dengan Λ adalah nilai *Wilk's lambda* pada persamaan (1). Nilai η_{Λ}^2 berada pada interval 0 dan 1, artinya semakin mendekati 1 berarti hubungan keeratan antar variabel respon dan variabel prediktor semakin erat.
(Rencher, 2002)

Pengujian Asumsi Regresi Linier Multivariat :

Uji Asumsi Residual Berdistribusi Normal Multivariat

Pemeriksaan residual yang harus dipenuhi dalam regresi linier multivariat adalah residual berdistribusi normal multivariat, pengujian dapat dilakukan dengan cara membuat Q-Q plot dari nilai d_i^2 (Johnson and Wichern, 2007).

Uji Hipotesis

Hipotesis

H_0 : Residual berdistribusi normal multivariat

H_1 : Residual tidak berdistribusi normal multivariat

Taraf Signifikan $\alpha = 5 \%$

Statistik Uji

$$r_Q = \frac{\sum_{j=1}^n \left(\hat{\epsilon}_{(i)} - \bar{\epsilon} \right) (q_{(i)} - \bar{q})}{\sqrt{\sum_{i=1}^n \left(\hat{\epsilon}_{(i)} - \bar{\epsilon} \right)^2 \sum_{i=1}^n (q_{(i)} - \bar{q})^2}}$$

Kriteria Uji:

Menolak H_0 pada taraf signifikansi α jika $r_Q < r$ tabel (*r correlation coefficient Test for Normality*).

(Johnson and Wichern, 2007)

Uji Homogenitas Matriks Varian Kovarian

Asumsi yang harus dipenuhi dalam regresi linier multivariat adalah residual memiliki matriks varian-kovarian yang homogen. Pengujian dilakukan dengan statistik uji Box's M.

Hipotesis

H_0 : $\Sigma_1 = \Sigma_2 = \dots = \Sigma_g$ (Matriks varian kovarian homogen)

H_1 : $\Sigma_r \neq \Sigma_s$ untuk paling sedikit dua matriks kovarian, dengan $r, s = 1, 2, \dots, g$

Taraf Signifikansi $\alpha = 5\%$

Statistik Uji

$$C = (1 - u) \left\{ \left[\sum_1 (n_1 - 1) \right] \ln |S_{\text{pooled}}| - \sum_1 [(n_1 - 1) \ln |S_1|] \right\}$$

Kriteria Uji

Menolak H_0 pada taraf signifikansi α jika $C > \chi_{p(p+1)(g-1)/2}^2(\alpha)$.

(Johnson and Wichern, 2007)

Uji Asumsi Residual Independen

Residual $\epsilon_1, \epsilon_2, \dots, \epsilon_q$ dikatakan bersifat saling bebas (independent) jika matriks korelasi antar residual membentuk matriks identitas. Pengujian kebebasan antar residual dapat dilakukan dengan uji *Bartlett Sphericity* sebagai berikut :

Hipotesis

H_0 : $P = I$ (Antar variabel respon bersifat independen)

H_1 : $P \neq I$ (Antar variabel respon tidak bersifat independen)

Taraf signifikansi $\alpha = 5 \%$

Statistik Uji

$$\chi_{\text{hitung}}^2 = - \left\{ n - 1 - \frac{2q + 5}{6} \right\} \ln |R|$$

Kriteria Uji

Menerima H_0 taraf signifikansi α jika $\chi^2_{hitung} \leq \chi^2_{\alpha; \frac{1}{2}q(q-1)}$ yang berarti residual bersifat independent.

(Morrison, 2005)

2.5 Influential Observation

Pada beberapa kasus seringkali terdapat satu atau lebih pengamatan yang menyebabkan perubahan signifikan pada kesesuaian fungsi regresi, pengamatan-pengamatan ini kemudian didefinisikan sebagai *Influential Observation* (Altunkaynak and Ekni, 2002). Jarak Cook tergeneralisasi digunakan untuk mengkaji pengaruh dari vektor pengamatan pada estimasi parameter. Pengukuran ini merupakan teknik yang berpengaruh untuk model regresi linier multivariat (Altunkaynak and Ekni, 2002).

2.6 Model Regresi Linier Multivariat yang Dibatasi

Misal $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}$ adalah model regresi linier multivariat dengan estimasi parameter untuk $\boldsymbol{\beta}$ adalah $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Model regresi linier yang dibatasi adalah: $\mathbf{T} = \mathbf{A}'\mathbf{Y}$ maka model regresi linier multivariat yang dibatasi adalah: $\mathbf{Y} = \mathbf{P}_A\mathbf{X}\boldsymbol{\beta} + \mathbf{E}$

(Altunkaynak and Ekni, 2002)

Beda antara estimasi parameter model regresi linier multivariat dengan regresi linier multivariat yang dibatasi adalah :

$$\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{\psi} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Z}(\mathbf{Z}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Z})^{-1}\mathbf{Z}'(\mathbf{I} - \mathbf{P}_X)\mathbf{Y}$$

(Altunkaynak and Ekni, 2002)

2.7 Jarak Cook Tergeneralisasi

Rumus Jarak Cook tergeneralisasi adalah :

$$DC_{\psi} = \frac{n-p}{p} \text{tr} \left\{ \mathbf{L}'(\mathbf{X}'\mathbf{X})\mathbf{L}(\mathbf{E}'\mathbf{E})^{-1} \right\}$$

Dalam hal ini jika nilai $DC_{\psi} > 1$ maka vektor – vektor pengamatan dapat dikatakan berpengaruh besar pada estimasi parameter.

(Altunkaynak and Ekni, 2002)

3. METODOLOGI PENELITIAN

3.1 Jenis dan Sumber Data

Data yang digunakan merupakan data sekunder dari hasil pendataan Badan Pusat Statistik (BPS) yaitu Statistik Pendidikan Provinsi Jawa Tengah Tahun 2010.

3.2 Variabel Penelitian

Variabel penelitian yang digunakan sebagai berikut :

a. Variabel Respon

Y_1 = Persentase Angka Partisipasi Kasar (APK) Jenjang Pendidikan SMP

Y_2 = Persentase Angka Partisipasi Sekolah (APS) Jenjang Pendidikan SMP

Y_3 = Persentase Angka Partisipasi Murni (APM) Jenjang Pendidikan SMP

b. Variabel Prediktor

X = Persentase Penduduk Berumur 10 Tahun ke Atas menurut Kabupaten atau Kota dan Pendidikan Tertinggi yang Ditamatkan (SMP)

3.3 Metode Analisis Data

Analisis yang digunakan untuk mengetahui *Influential Observation* pada model regresi linier multivariat sebagai berikut :

1. Uji Korelasi Variabel Respon
2. Model Regresi Linier Multivariat Indikator Pendidikan Provinsi Jawa Tengah Tahun 2010
3. Pengujian Hipotesis dan Estimasi Parameter Model Regresi Linier Multivariat
4. Pengujian Asumsi Regresi Linier Multivariat
5. Menurut Altunkaynak and Ekni (2002), langkah - langkah pendeteksian *Influential Observation* menggunakan jarak Cook tergeneralisasi sebagai berikut :

a. Menghitung matriks $K = (I - P_X)$, dengan I adalah matriks identitas dan $P_X = X(X'X)^{-1}X'$

b. Menghitung matriks residual $E = (I - P_X)Y = (Y - X\hat{\beta})$

c. Menghitung beda estimasi parameter model regresi multivariat dengan model regresi linier multivariat yang dibatasi.

$$L = \hat{\beta} - \hat{\beta}_\psi = (X'X)^{-1}X'I^D(K_D^D)^{-1}I_D(Y - X\hat{\beta})$$

d. Menghitung nilai jarak cook tergeneralisasi

$$DC_\psi = \frac{n-p}{p} \text{trace}\{(L'X'XL(E'E)^{-1})\}$$

e. Jika nilai $DC_\psi > 1$, maka terdapat *Influential Observation*.

f. Menghapus data yang terdapat *Influential Observation*.

g. Menganalisis perubahan koefisien regresi setelah data yang terdapat *influential observation* dihapus.

4. PEMBAHASAN

4.1 Analisis Hubungan Variabel Respon

Adanya korelasi antara variabel respon yang signifikan dapat dikatakan bahwa ketiga variabel respon yaitu APK, APS, dan APM memiliki hubungan.

Tabel 1 Hubungan Antar Variabel Respon

Variabel Respon	APK (Y ₁)	APS (Y ₂)	APM (Y ₃)
APK (Y ₁)	1	0.736	0.811
APS (Y ₂)	0.736	1	0.848
APM (Y ₃)	0.811	0.848	1

Nilai korelasi antara variabel respon dapat dilihat pada Tabel 1, nilai korelasi Angka Partisipasi Kasar (APK) dan Angka Partisipasi Sekolah (APS) sebesar 0.736 hal ini dapat diartikan bahwa antara Angka Partisipasi Kasar dan Angka Partisipasi Sekolah berbanding lurus, artinya jika Angka Partisipasi Kasar tinggi maka Angka Partisipasi Sekolah tinggi. Nilai korelasi antara Angka Partisipasi Kasar dengan Angka Partisipasi Murni sebesar 0.811, dapat diartikan bahwa Angka Partisipasi Kasar dan Angka Partisipasi Murni berbanding lurus, artinya jika Angka Partisipasi Kasar tinggi maka Angka Partisipasi Murni tinggi. Nilai korelasi antara Angka Partisipasi Sekolah dan Angka Partisipasi Murni sebesar 0.848, korelasi ini berbanding lurus. Jika Angka Partisipasi Sekolah tinggi maka Angka Partisipasi Murni tinggi. Dapat disimpulkan bahwa ketiga variabel respon saling berkorelasi.

4.2 Model Regresi Linier Multivariat Indikator Pendidikan di Jawa Tengah Sebelum Pendeteksian *Influential Observation*

Model regresi linier multivariat untuk masing – masing variabel respon yaitu Angka Partisipasi Kasar (Y₁), Angka Partisipasi Sekolah (Y₂), dan Angka Partisipasi Murni (Y₃) dan variabel prediktor yaitu persentase penduduk usia 10 tahun ke atas yang menamatkan jenjang SMP (X) sebelum dideteksi adanya *Influential Observation* sebagai berikut :

$$\hat{Y} = X\hat{\beta} \text{ dengan } \hat{\beta} = \begin{bmatrix} 0.5700 & 0.5496 & 0.5131 \\ 1.3467 & 1.6980 & 1.0357 \end{bmatrix}$$

Interpretasi model tersebut adalah, jika persentase penduduk usia diatas 10 tahun berdasarkan pendidikan tertinggi yang ditamatkan yaitu SMP bertambah satu persen akan berdampak pada meningkatnya Angka Partisipasi Kasar berdasarkan jenjang pendidikan SMP sebesar 1.3467 %, meningkatnya Angka Partisipasi Sekolah berdasarkan jenjang pendidikan SMP sebesar 1.6980 % dan meningkatnya Angka Partisipasi Murni berdasarkan jenjang pendidikan SMP sebesar 1.0357 %.

Pengujian Regresi Linier Multivariat Secara Keseluruhan

Hipotesis

$H_0 : \beta_{11}=\beta_{21}=\beta_{31}=0$ (Parameter tidak signifikan secara bersama-sama terhadap model)

$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$ (Parameter signifikan secara bersama-sama terhadap model)

Statistik Uji

$$\Lambda = \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - n\bar{y}\bar{y}'|} = 0.0042$$

Hasil pengujian diketahui nilai *Wilk's Lambda* adalah 0.0042 dan nilai *Wilk's Lambda* tabel yaitu $\Lambda_{0.05;3;1;33} = 0.7768$. Menolak H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai *Wilk's Lambda* < nilai *Wilk's Lambda* tabel, maka dapat disimpulkan bahwa secara multivariat variabel prediktor dalam model berpengaruh terhadap variabel respon yaitu APK, APS, dan APM.

Pengujian Signifikansi Koefisien Parameter pada Model Regresi Linier Multivariat

Hipotesis

$H_0 : \beta_{11}=\beta_{21}=\beta_{31}=0$ (Koefisien dari Parameter untuk Variabel X tidak signifikan terhadap model)

$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$ (Koefisien dari Parameter untuk Variabel X signifikan terhadap model)

Statistik Uji

$$\Lambda_x = \frac{\Lambda_f}{\Lambda_r} = 0.506$$

Nilai $\Lambda_x = 0.664$, dan nilai *Wilk's Lambda* tabel yaitu $\Lambda_{0.05;3;3;33} = 0.5894$. Menolak H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai $\Lambda_x < \Lambda_{0.05;3;3;33} = 0.5894$, maka dapat disimpulkan bahwa variabel X yaitu persentase penduduk berusia diatas 10 tahun yang menamatkan jenjang pendidikan SMP berpengaruh terhadap variabel respon yaitu APK, APS dan APM.

Pengujian Asumsi Regresi Linier Multivariat

Uji Residual Berdistribusi Normal Multivariat

Pengujian secara uji visual (grafik), sebaran nilai residual untuk plot *Chi Square* berada di sekitar garis lurus, yang berarti bahwa residual berdistribusi normal multivariat. Selanjutnya digunakan pengujian

hipotesis sebagai berikut :

$H_0 : \text{Residual berdistribusi normal multivariat}$

$H_1 : \text{Residual tidak berdistribusi normal multivariat}$

Menerima H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai $r_Q = 0.972 > r_{0.05;35} = 0.9682$, residual berdistribusi normal multivariat. Asumsi residual berdistribusi normal multivariat terpenuhi.

Uji Asumsi Homogenitas Matriks Varian Kovarian

Asumsi selanjutnya yang harus dipenuhi dalam pemodelan secara multivariat adalah matriks varians–kovarian residual homogen. Pengujian dilakukan terhadap nilai dari residual dengan hipotesis sebagai berikut:

$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ (matriks varian kovarian homogen)

$H_1 : \Sigma_r \neq \Sigma_s$ (matriks varian kovarian tidak homogen)

Menerima H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai statistik uji Box's M adalah $1.253 < \chi^2_{0.05;6} = 12.59$ dan $P \text{ value} = 0.985 > \alpha = 0.05$, asumsi matriks varian kovarian residual bersifat homogen terpenuhi.

Residual Saling Bebas (Independen)

Selain asumsi matrik varian-kovarian homogen, asumsi residual saling bebas juga harus dipenuhi. Pengujian asumsi ini dilakukan dengan uji *Bartlett Sphericity*. Bentuk hipotesisnya sebagai berikut :
 $H_0 : \mathbf{P} = \mathbf{I}$ (Residual bersifat independen)

$H_1 : \mathbf{P} \neq \mathbf{I}$ (Residual bersifat dependen)

Menerima H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai *Chi Square test* = $4.722 < \chi^2_{0.05;3} = 7.81$ dan $P \text{ value} = 0.193 > \alpha = 0.05$. Asumsi residual bersifat independen terpenuhi.

4.3 Pendeteksian *Influential Observation* Pada Indikator Pendidikan di Jawa Tengah

Tabel 2 Nilai Jarak Cook Tergeneralisasi untuk Masing-Masing Pengamatan

D_i	DC_i	D_i	DC_i	D_i	DC_i	D_i	DC_i	D_i	DC_i
1	0.0184	8	0.0667	15	0.3433	22	0.1045	29	0.4054
2	0.1361	9	0.2409	16	0.7136	23	0.0826	30	0.1406
3	0.0521	10	0.0848	17	0.0781	24	0.0195	31	0.0283
4	0.2194	11	0.1806	18	0.1199	25	0.0136	32	0.1422
5	0.0820	12	0.1810	19	0.1073	26	0.2022	33	0.1418
6	0.0995	13	0.0332	20	0.6416	27	0.0515	34	0.0410
7	1.2895	14	3.4332	21	0.5557	28	0.0743	35	0.1530

Hasil perhitungan jarak Cook tergeneralisasi untuk masing-masing pengamatan ditunjukkan pada Tabel 2 terdapat 2 pengamatan yang terpisah jauh dari sekumpulan data. Pengamatan ke 7 dan 14 merupakan *Influential Observation* karena nilai jarak Cook tergeneralisasi untuk pengamatan tersebut lebih besar dari 1 yaitu $DC_{(7)}=1.2985$, dan $DC_{(14)}=3.4332$.

4.4 Analisis Pengaruh Penghapusan Pengamatan Terhadap Model Regresi Linier Multivariat Model Regresi Linier Multivariat dengan Penghapusan Pengamatan ke 7 dan 14

Model regresi linier multivariat untuk masing – masing variabel respon Angka Partisipasi Kasar (Y_1), Angka Partisipasi Sekolah (Y_2), dan Angka Partisi Murni (Y_3) dan variabel prediktor yaitu persentase penduduk usia diatas 10 tahun yang tamat SMP (X) dengan penghapusan pengamatan ke 7 dan 14 sebagai berikut :

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \text{ dengan } \hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.4958 & 0.5682 & 0.4568 \\ 1.7849 & 1.6275 & 1.3712 \end{bmatrix}$$

Interpretasi model tersebut sebagai berikut, jika persentase penduduk usia diatas 10 tahun berdasarkan pendidikan tertinggi yang ditamatkan yaitu SMP bertambah satu persen akan berdampak pada meningkatnya Angka Partisipasi Kasar berdasarkan jenjang pendidikan SMP sebesar 1.7849 % , meningkatnya Angka Partisipasi Sekolah berdasarkan jenjang pendidikan SMP sebesar 1.6275 % dan meningkatnya Angka Partisipasi Murni berdasarkan jenjang pendidikan SMP sebesar 1.3712 %.

Pengujian Regresi Linier Multivariat Secara Keseluruhan dengan Penghapusan Pengamatan ke 7 dan 14

Hipotesis

$H_0 : \beta_{11}=\beta_{21}=\beta_{31}=0$ (Parameter tidak signifikan secara bersama-sama terhadap model)

$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$ (Parameter signifikan secara bersama-sama terhadap model)

$$\Lambda = \frac{|\mathbf{Y}'\mathbf{Y} - \hat{\mathbf{B}}'\mathbf{X}'\mathbf{Y}|}{|\mathbf{Y}'\mathbf{Y} - n\bar{y}\bar{y}'|} = 0.0037$$

Hasil pengujian diketahui nilai *Wilk's Lambda* adalah 0.0039, dan nilai *Wilk's Lambda* tabel yaitu $\Lambda_{0.05;3;1;31} = 0.7656$. Menolak H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai *Wilk's Lambda* < nilai *Wilk's Lambda* tabel, maka dapat disimpulkan bahwa secara multivariat variabel prediktor dalam model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14 berpengaruh terhadap variabel respon yaitu APK, APS, dan APM.

Pengujian Signifikansi Koefisien Parameter pada Model Regresi Linier Multivariat dengan Penghapusan Pengamatan ke 7 dan 14

Hipotesis

$H_0 : \beta_{11}=\beta_{21}=\beta_{31}=0$ (Koefisien dari Parameter untuk Variabel X_1 tidak signifikan terhadap model)

$H_1 : \text{Paling sedikit ada satu } \beta_{jk} \neq 0$ (Koefisien dari Parameter untuk Variabel X_1 signifikan terhadap model)

Statistik Uji

$$\Lambda_x = \frac{\Lambda_f}{\Lambda_r} = 0.549$$

Nilai $\Lambda_x = 0.741$, dan nilai *Wilk's Lambda* tabel yaitu $\Lambda_{0.05;3;3;31} = 0.5718$, menolak H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai $\Lambda_x < \Lambda_{0.05;3;3;31} = 0.5718$, maka dapat disimpulkan bahwa variabel X_1 yaitu persentase penduduk usia diatas 10 tahun yang menamatkan jenjang pendidikan SMP berpengaruh terhadap variabel respon yaitu APK, APS, dan APM pada model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14.

Pengujian Asumsi Regresi Linier Multivariat dengan Penghapusan Pengamatan ke 7 dan 14

Uji Residual Berdistribusi Normal Multivariat

Pengujian secara uji visual (grafik) diperoleh sebaran nilai residual untuk plot *Chi Square* dengan penghapusan pengamatan ke 7 dan 14 berada di sekitar garis lurus, yang berarti bahwa residual berdistribusi normal multivariat. Selanjutnya digunakan pengujian hipotesis sebagai berikut :

H_0 : Residual berdistribusi normal multivariat

H_1 : Residual tidak berdistribusi normal multivariat

Menerima H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai $r_Q = 0.9753 > r_{0.05;32} = 0.9682$. Asumsi residual berdistribusi normal multivariat untuk model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14 terpenuhi.

Uji Asumsi Homogenitas Matriks Varian Kovarian

Uji homogenitas untuk matriks varian kovarian pada model regresi linier multivariat penghapusan pengamatan ke 7 dan 14 sebagai berikut:

Hipotesis

$H_0 : \Sigma_1 = \Sigma_2 = \Sigma_3$ (matriks varian kovarian homogen)

$H_1 : \Sigma_r \neq \Sigma_s$ (matriks varian kovarian tidak homogen)

Menerima H_0 pada taraf signifikansi $\alpha=5\%$ karena nilai Box M test = 1.311 < $\chi^2_{0.05;6} = 12.59$ dan P value = 0.983 > $\alpha = 0.05$. Asumsi matriks varian kovarian residual bersifat homogen untuk model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14 terpenuhi.

Residual Saling Bebas (Independen)

Uji asumsi residual saling bebas untuk model regresi linier multivariat penghapusan pengamatan ke 7 dan 14 sebagai berikut :

Hipotesis

$$H_0 : P = I \text{ (Residual bersifat independen)}$$

$$H_1 : P \neq I \text{ (Residual bersifat dependen)}$$

Menerima H_0 pada taraf signifikansi $\alpha = 5\%$ karena nilai *Chi Square test* = 4.133 < $\chi^2_{0.05;3} = 7.81$ dan P value = 0.247 > $\alpha = 0.05$. Asumsi residual bersifat independen untuk model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14 terpenuhi.

4.5 Hubungan Antara Variabel Respon dan Variabel Prediktor

Pengaruh penghapusan pengamatan ke 7 dan 14 selain berdampak pada perubahan koefisien regresi, juga berdampak pada perubahan besarnya hubungan keeratan antara variabel respon dengan variabel prediktor. Ukuran keeratan menggunakan *Wilk's Lambda* yaitu :

$$\eta^2_{\Lambda} = 1 - \Lambda$$

Tabel 3 Perbandingan Hubungan Keeratan Model Regresi

Model Regresi	Nilai <i>Wilk's Lambda</i> (η^2_{Λ})
Tanpa penghapusan	0.9958
Penghapusan Pengamatan ke 7 dan 14	0.9963

Berdasarkan Tabel 3 perbandingan hubungan keeratan antara variabel respon dengan variabel prediktor sebelum dan sesudah penghapusan data ke 7 dan 14 tidak mengalami perubahan yang signifikan.

5. KESIMPULAN

Model regresi linier multivariat dengan penghapusan pengamatan ke 7 dan 14 sebagai berikut :

$$\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}} \text{ dengan } \hat{\boldsymbol{\beta}} = \begin{bmatrix} 0.4958 & 0.5682 & 0.4568 \\ 1.7849 & 1.6275 & 1.3712 \end{bmatrix}$$

Berdasarkan model regresi linier multivariat dapat dijelaskan bahwa jika persentase penduduk usia diatas 10 tahun berdasarkan pendidikan tertinggi yang ditamatkan yaitu SMP bertambah satu persen akan berdampak pada meningkatnya Angka Partisipasi Kasar berdasarkan jenjang pendidikan SMP sebesar 1.7849 % , meningkatnya Angka Partisipasi Sekolah berdasarkan jenjang pendidikan SMP sebesar 1.6275 % dan meningkatnya Angka Partisipasi Murni berdasarkan jenjang pendidikan SMP sebesar 1.3712 %. Besarnya hubungan keeratan antara variabel respon dan prediktor diperoleh nilai $\eta^2_{\Lambda} = 1 - 0.0037 = 0.9963$, dapat dikatakan bahwa model dapat menjelaskan informasi data sebesar 99.63 %. Penghapusan kedua pengamatan secara bersama-sama yang termasuk *Influential Observation* pada model regresi linier multivariat indikator pendidikan di Jawa Tengah hanya berpengaruh terhadap perubahan koefisien regresi saja dan tidak berdampak besar pada hubungan keeratan antara variabel respon dan variabel prediktor secara multivariat.

DAFTAR PUSTAKA

- Altunkaynak, B., and Ekni, M. 2002. Detection of Influential Observation Vectors for Multivariate Linear Regression. *Hacettepe Journal of Mathematics and Statistics* Vol. 31:139-151.
- Badan Pusat Statistik Jawa Tengah. 2010. *Statistik Pendidikan Provinsi Jawa Tengah: Hasil Susenas 2010*. Semarang: Badan Pusat Statistika.
- Johnson, R.A and D.W Wichern. 2007. *Applied Multivariate Statistical Analysis : Sixth Edition*. New Jersey : Prentice Hall International Inc.
- Morrison, D.F. 2005. *Multivariate Statistical Methods*, Fourth Edition, The Wharton School University of Pennsylvania.
- Rencher, C.A. 2002. *Methods of Multivariate Analysis. Second Edition*. New York: Jhon Wiley & Sons.
- Timm, H.N. 2002. *Applied Multivariate Analysis*. New York: Springer-Verlag.